What Fraction of Sun-like Stars have Planets?

Charles H. Lineweaver & Daniel Grether
University of New South Wales
charley@bat.phys.unsw.edu.au

ABSTRACT

The radial velocities of $\sim$ 1800 nearby Sun-like stars are currently being monitored by eight high-sensitivity Doppler exoplanet surveys. Approximately 90 of these stars have been found to host exoplanets massive enough to be detectable. Thus at least $\sim$ 5% of target stars possess planets. If we limit our analysis to target stars that have been monitored the longest ($\sim$ 15 years), $\sim$ 11% possess planets. If we limit our analysis to stars monitored the longest and whose low surface activity allow the most precise velocity measurements, $\sim$ 25% possess planets. By identifying trends of the exoplanet mass and period distributions in a sub-sample of exoplanets less-biased by selection effects, and linearly extrapolating these trends into regions of parameter space that have not yet been completely sampled, we find at least $\sim$ 9% of Sun-like stars have planets in the mass and orbital period ranges $M\sin\, i > 0.3 M_{Jupiter}$ and $P < 13$ years, and at least $\sim$ 22% have planets in the larger range $M\sin\, i > 0.1 M_{Jupiter}$ and $P < 60$ years. Even this larger area of the log mass - log period plane is less than 20% of the area occupied by our planetary system, suggesting that this estimate is still a lower limit to the true fraction of Sun-like stars with planets, which may be as large as $\sim$ 100%.

Subject headings: Planetology, Planetary systems, Extrasolar Planets

1. Introduction

With increasingly sensitive instruments exoplanet hunters have detected more than 100 exoplanets. The focus of these pioneering efforts has been to find and describe new exoplanets. As more exoplanets have been found, the question: 'What fraction of stars have planets?' has been looked at periodically. Estimates of the fraction of stars with planets can be simply calculated from the raw numbers of exoplanet hosts divided by the number of monitored stars. For example, Marcy & Butler (2000) report "5% harbor companions of 0.5 to $8 M_{Jupiter}$ within 3 AU". Estimates can also be based on high precision Doppler targets in a single survey. Fischer *et al.* (2003) report a fraction of 15% for high precision Doppler targets in the original Lick sample. Estimates can also be based on a semi-empirical analysis of exoplanet data (Tabachnik and Tremaine 2002).

In this work we use a semi-empirical method, staying as close to the exoplanet data as possible. We use the growth and current levels of exoplanet detection to verify that sensitive and long duration surveys have been finding more planets in a predictable way. This is no surprise. However, quantitatively following the consistent increase of the lower limit to the fraction of stars with detected planets is an important new way to substantiate both current and future estimates of this fraction. Because of the growing importance of the question 'What fraction of stars have planets?' and our increasing ability to answer this question, such a closer scrutiny of (i) the assumptions used to arrive at the answer, (ii) the parameter space in which they are valid and (iii) the associated error bars, is timely.

If the Sun were among the target stars of the Doppler surveys, it would be another few years before the presence of Jupiter (the most easily detected feature of our planetary system) could be confidently detected. However, we are beginning to be able to answer the question: How typical is Jupiter? In Lineweaver & Grether (2002) we found that extrapolations of trends found in a less-biased sub-sample of exoplanet data indicate that Jupiter is a typical massive planet in the sense that it lies in the most densely occupied region of the log mass - log period plane. In Lineweaver *et al.* (2003) we updated our analysis to include exoplanets detected between January and August 2002 and found more support for this idea. However, in this previous work we made no effort to determine the absolute frequency of Jupiters or to answer the question: What fraction of stars have planets? That is the main focus of this paper.

In Section 2 we present the exoplanet data set (Fig. 1). We analyze the target lists and detections and show how the fraction of stars with detected planets has increased over time (Figs. 2 & 3). In Section 3 we quantify how the fraction of stars with planets depends on the precision with which the radial velocity of individual stars can be measured (Fig. 4). In Section 4 we quantify trends in exoplanet mass and period by linear and power-law fits to histograms of a less-biased sub-sample of exoplanets (Fig. 5). Based on these trends, we extrapolate into larger regions of parameter space and give estimates for the fraction of stars with planets in these regions. In Section 5 we summarize, discuss and compare our results with previous work. Our approach is most similar and complimentary to the work of Fischer *et al.* (2003) and Tabachnik & Tremaine (2002).

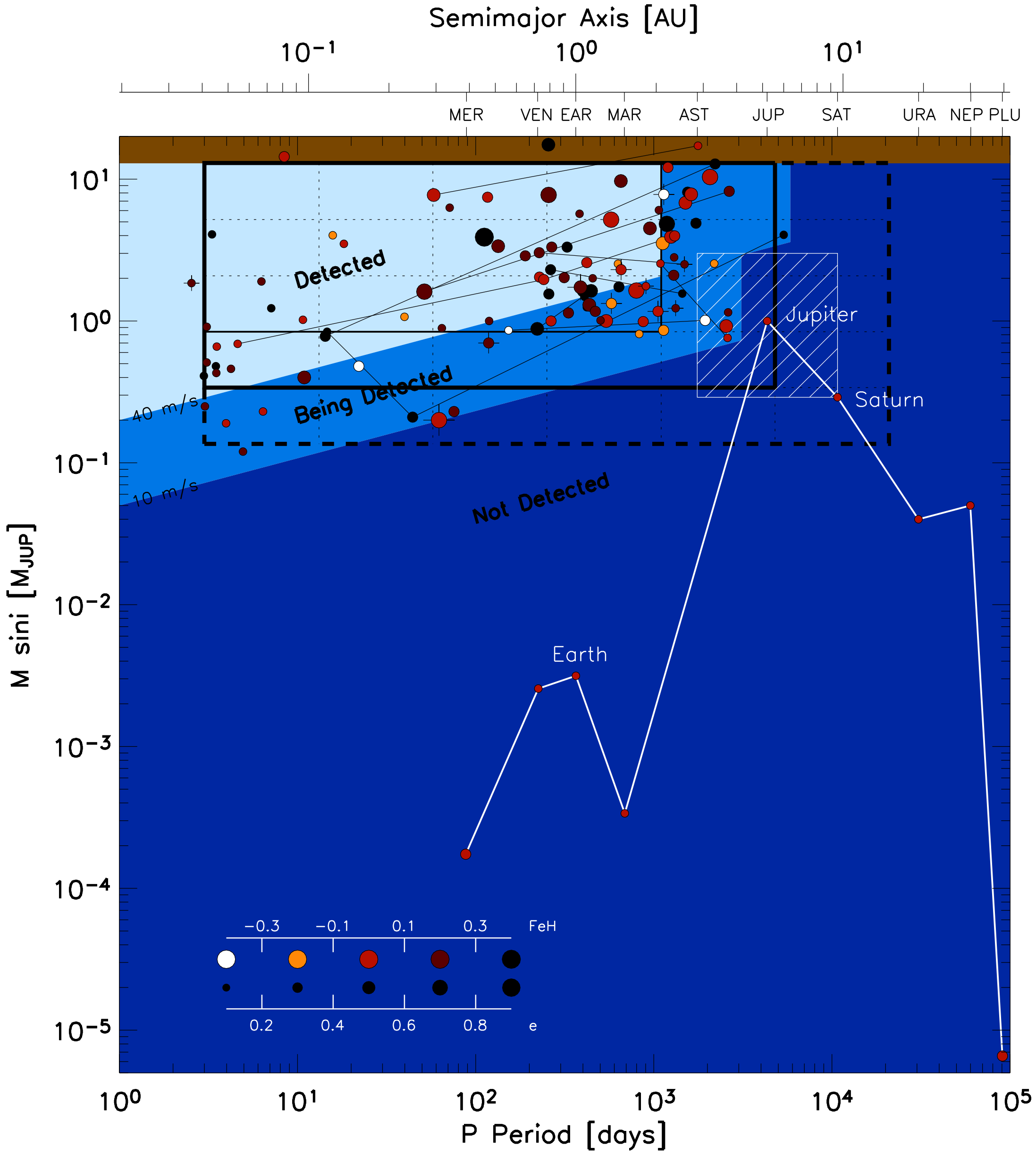

Fig. 1.— Our Solar System compared to the 106 exoplanets detected by eight high-sensitivity Doppler surveys. The three regions labeled "Detected", "Being Detected" and "Not Detected" indicate the selection effects in the log mass - log period plane due to a limited time of observation and limited radial velocity sensitivity (Section 2.1). The rectangular region circumscribed by the thin solid line, fits almost entirely into the "Detected" region, and contains what we call the less-biased sub-sample. This sub-sample is the basis of the trends in mass and period identified in Fig. 5. Metallicity of the host stars are indicated by the shade of gray of the point while eccentricity of the exoplanet orbits are indicated by point size (see key in lower left). Thin lines connect planets orbiting the same host star. The cross-hatched square is our Jupiter-like region defined by $M_{Saturn} \leq M\sin i \leq 3M_{Jupiter}$ and $P_{asteroids} \leq P \leq P_{Saturn}$.

2. Exoplanet Data

2.1. Mass-period Distribution

Figure 1 displays the masses and periods of the 106 exoplanets detected as of June 2003 by the eight high precision Doppler surveys analyzed in this paper. The region in the upper left labeled "Detected" is our estimate of the region in which the Doppler technique has detected virtually all the exoplanets with periods less than three years orbiting target stars that have been monitored for at least three years. This region is bounded by three days on the left, three years on the right, $13M_{Jupiter}$ on the top and at the bottom by a radial velocity of 40 m/s induced in a solar-mass host star. The largest observed exoplanet period and the smallest observed radial velocity induced by a detected exoplanet are used to define the boundary between the "Being Detected" and "Not Detected" regions. The discontinuity of the "Being Detected" region near Jupiter is due to the increased sensitivity of the original Lick survey at the end of 1994. No more exoplanets should be detected in the "Detected" region unless new stars are added to the target lists. Thus, all the exoplanets marked with a '+' (detected since August 2002) should fall in, or near, the "Being Detected" region. Of the 11 new detections since August 2002, 8 fall in the "Being Detected" region, 2 fall just inside the "Detected" region while one has $P < 3$ days – it was not being monitored with sufficient phase coverage for detection until recently.

We define a less-biased sample of planet hosts as the 49 hosts to the planets within the rectangular region circumscribed by a thin solid line (3 days $< P <$ 3 years and $0.84 < \frac{M\sin i}{M_{Jupiter}} < 13$) in Fig. 1. This rectangle is predominantly in the "Detected" region. We will use this less-biased sample as the basis for our extrapolations (Sec. 4). As a completeness correction for the lower right corner of this rectangle being in the 'Being Detected' region, we add 4 planets as described in (Lineweaver *et al.* 2003) giving us a corrected less-biased sample of $53(= 49 + 4)$ planets.

2.2. Increasing Fraction as a function of monitoring duration

The exoplanets plotted in Fig. 1 are the combined detections of eight Doppler surveys currently monitoring the radial velocities of ≈ 1812 nearby FGK stars (Tables 3 & 4). The top panel of Figure 2 shows how the number of these target stars has increased over the past 16 years. 94 stars have been found to host 106 exoplanets. Of these 94, 92 fall within our selection criteria of Sun-like stars (= FGK class IV or V) with planets ($M\sin i < 13M_{Jupiter}$). Six known exoplanet hosts were not included in this analysis because they were found in the context of surveys whose search strategies and sensitivities cannot easily be combined with results from the 8 sensitive Doppler surveys analyzed here.

In Fig. 2, the date-of-first-monitoring and the discovery date for the detected exoplanets were largely obtained from the literature and press releases (Table 3). In some cases the date-of-first-monitoring was estimated from the first point on the planet host's velocity curve. The ramp-up time needed to start observing all the stars in a survey's target list was estimated from the distribution of the date-of-first-montitoring of a survey's detected exoplanets. That is, the ramp up time needed to start observing the detected hosts was used to estimate the ramp-up time needed to start observing all the stars on a survey's target list. The exceptions to this were the time dependence of the original Lick target list obtained from Cumming *et al.* (1999) and the time dependence of the Keck target list (FGK) estimated from the FGKM histogram of Cumming *et al.* (2003).

The date-of-first-monitoring binning (middle-gray histogram in the top panel of Fig. 2) shows that the fraction of target stars hosting planets has a maximum of $\approx 11\%$ from the longest monitored targets and then decreases as we average in target stars that have been monitored for shorter times. The two binning conventions used (date-of-detection and date-of-first-monitoring) are equivalent today at the current value of $5\% (\approx 92/1812)$.

Although this 11% estimate is based on the results from ~ 85 target stars of the two longest running surveys: Lick and McDonald, it is consistent with the increasing fraction based on the monitoring duration of all target stars. This is shown in Fig. 3. We fit a curve to this data normalized at a duration of 15 years to the weighted average fraction of the two longest duration bins (Sec. 4.3). The extrapolation of this curve (based on current sensitivity) to monitoring durations of 30 years (approximate period of Saturn) yields a fraction of $\sim 15\%$. This extrapolation corresponds to extending the white "Detected" region in Fig. 1 to the upper right, above the diagonal $K = 40$ m/s line.

In the lower panel of Fig. 3, we increase the fraction for a given group of stars, not when a planet is reported, but at the duration equal to the period of the newly detected planet. For example, recent detections of short period planets increase the fraction at a duration corresponding to the short period of the newly detected planet. This smooths over artificial delays associated, for example, with not analyzing the data for the first 8 years of observations, and enables us to trace with dotted lines the increasing fraction for each group of target stars.

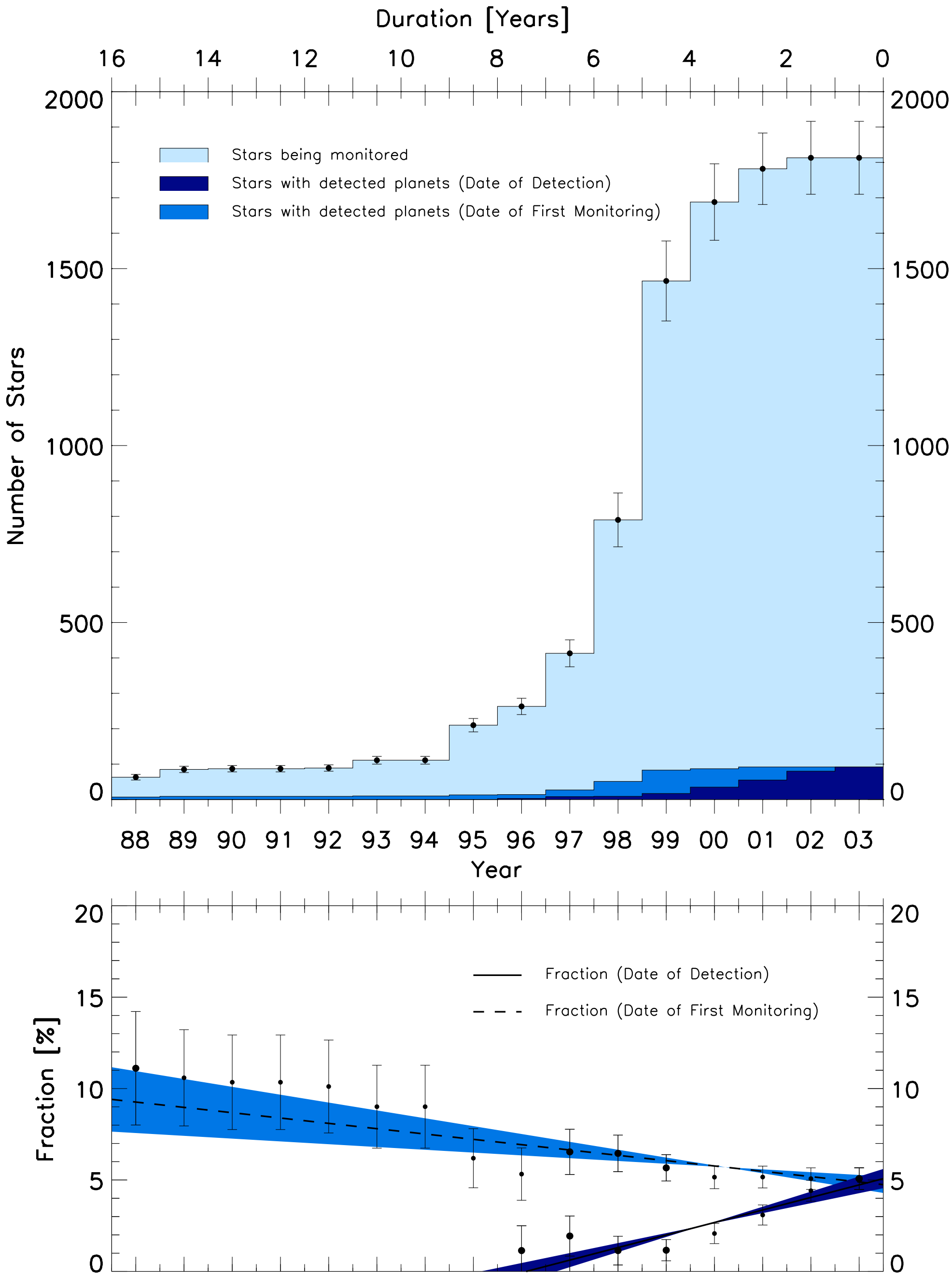

Fig. 2.— The fraction of stars with planets. TOP: Histograms of the increasing number of target stars and the increasing number of them found to be hosting at least one exoplanet. The host stars are binned in two ways. In the darkest histogram they are binned at the date of detection of the first exoplanet found orbiting the star. In the middle-gray histogram they are binned at the date of first monitoring of the star. Notice the small number of target stars until the discovery of the first planet (Mayor & Queloz 1995). BOTTOM: By taking the ratio, in the top panel, of the darker histograms to the lightest one, we obtain the fraction of target stars hosting at least one planet. The two binning conventions lead to different results. Using the date-of-detection binning (darkest histogram) yields the intuitive result that the fraction of host stars starts at zero in 1995 and climbs monotonically until its current value of $5 \pm 1\%$ ($\approx 92/1812$). The date-of-first-monitoring binning starts on the far left at $11 \pm 3\%$ from the bin of target stars that has been monitored the longest. The fraction decreases as we average in more stars that have been monitored for shorter durations. The dashed line is the result of a linear fit to a sparse, and therefore more independent, sample of the non-independent points (data points used have a larger point size). This line yields $9 \pm 2\%$ on the far left in good agreement with the single data point in the left-most bin. The last bin on the right is the same in both binning methods and includes all Sun-like target stars. Thus, $\approx 5\%$ is an average fraction from target stars that have been monitored for times varying between 0 and 16 years, while $\approx 11\%$ is the fraction from target stars that have been monitored the longest (≈ 15 yrs). The non-trivial task of estimating the total number of stars being monitored is described in Appendix A and summarized in Table 4.

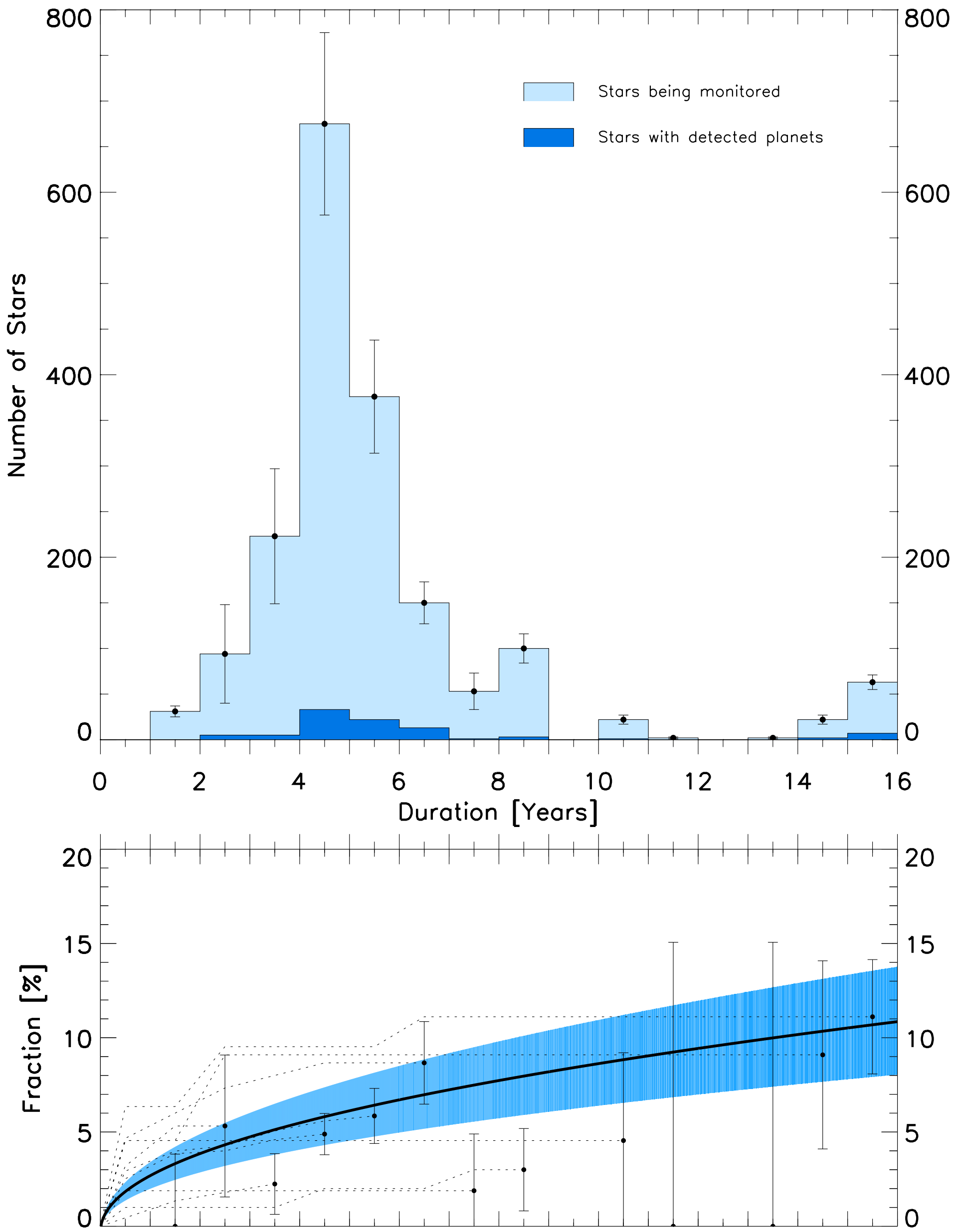

Fig. 3.— Number of targets as a function of how many years they have been monitored. Most targets have only been monitored for ∼ 5 years. The longer a group of target stars has been monitored, the larger the fraction of planet-possessing targets. TOP: Notice the small number of target stars (∼ 90) in the original Doppler surveys (bins on far right) and the ∼ 2 years it took to start monitoring all of them. These ∼ 90 (∼ 5% of the stars currently being monitored) are the only target stars that have been monitored long enough to begin to detect exoplanets with Jupiter-like masses in Jupiter-like (∼ 12 year) orbital periods. This sample in the two longest duration bins has resulted in 9 detections from 85 targets for a fraction of ≈ 11%. It will be another five years before a substantially larger fraction of target stars have been monitored long enough to detect such planets. BOTTOM: The fractions plotted here are the ratios of the histograms in the upper panel as in the previous figure. This panel shows the effect of longer monitoring duration on the observed fraction of stars with planets. The fraction of target stars with detected planets increases with duration. This is distinct from the cumulative fractions shown in the lower panel of Fig. 2. However, the right-most bin here is the same as the left-most bin there. Calculation of the fraction of stars hosting exoplanets (the curve in the lower panel with the light gray 68% confidence region) is described in Sec. 4.3. The scatter of the data points around the smooth curve is due to small number statistics, varying instrumental sensitivities and variation in observational phase coverage. For example, the two points at durations of 7.5 and 8.5 years lie below the curve, due to the lower average sensitivity $\gtrsim 10$ m/s of AFOE and Elodie surveys in these bins. The high point at 6.5 years is the start of the sensitive Keck survey. Thus, this plot also shows the effect of different instrumental sensitivities on the fraction of stars hosting planets.

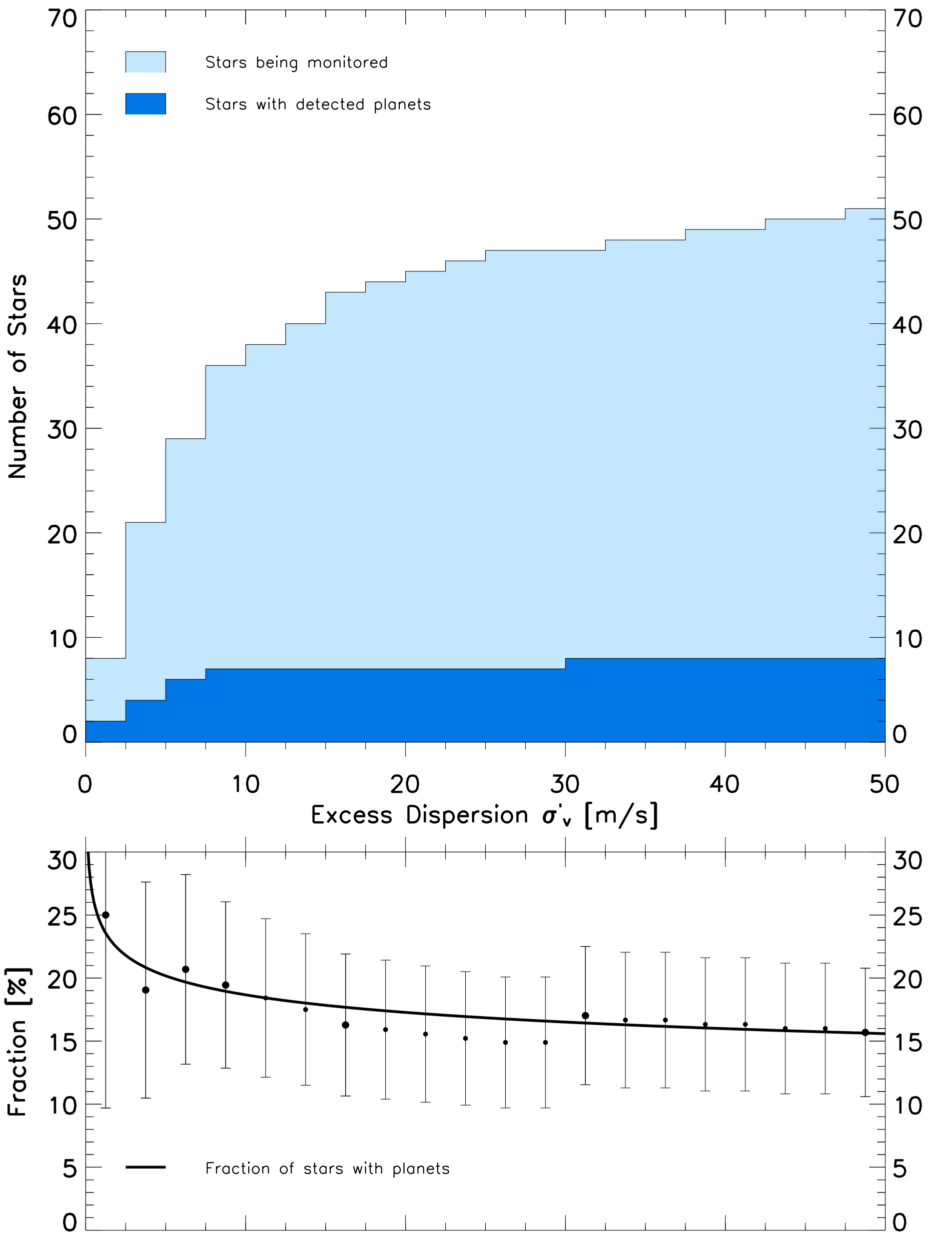

Fig. 4.— When stellar activity is low, high measurement precision is possible and a higher fraction of targets are found to host planets. TOP: The number of targets in the original Lick survey that have been monitored for a duration of 14-16 years and that have an excess radial velocity dispersion σ_v' less than the value on the x axis. σ_v' is a measure of how precisely one can determine the radial velocity of the targets. The number of detections from these targets is also shown. The ratio (BOTTOM) gives the fraction of stars with planets as a function of σ_v' threshold. The σ_v' of the Lick target stars have been estimated from the stellar rotational periods tabulated by Cumming *et al.* (1999) using the relations $\sigma_v' = 10 * (12/P_{rot})^{1.1}$ for G and K stars and $\sigma_v' = 10 * (10/P_{rot})^{1.3}$ for F stars (Saar *et al.* 1998). The curve is a power law ($df/d\sigma_v' \propto {\sigma_v'}^{\gamma}$) fit to a sparse sample of these non-independent points (larger point size). Thus, as we select for higher Doppler precision in the Lick sample, the fraction of targets possessing planets increases from 15% to 25% for the stars with the highest precision ($\sigma_v' < 2.5$ m/s). Small number statistics increases the error bars as the decreasing σ_v' threshold reduces the number of targets. In this cumulative plot the error bars are highly correlated.

3. High Doppler Precision Targets

Instrumental sensitivity is an important limit on a Doppler survey's ability to detect planets. However, the level of stellar activity on a star's surface is also important. Using the Doppler technique, planets are more easily detected around slowly rotating stars with low level chromospheric activity, little granulation or convective inhomogeneities and few time-dependent surface features. To reduce these problems, some target lists have been selected for high Doppler precision (low stellar activity) by excluding targets with high values of projected rotational velocity $v\sin i$ or chromospheric emission ratio R'_{HK} (Table 3).

By selecting target stars monitored the longest, the detected fraction of Sun-like stars possessing planets increased from 5% to 11%. By selecting from the target stars monitored the longest, the stars with the highest Doppler precision, the fraction increases still further. Fischer *et al.* (2003) analyzed a group of high Doppler precision target stars in the original Lick sample and found that 15% possessed planets. We extend this idea in Fig. 4 by plotting the fraction of target stars with planets as a function of their stellar activity as measured by their excess radial velocity dispersion σ'_v (Saar *et al.* 1998). We find that this selection increases the fraction from 15% on the far right of Fig. 4 to $\sim$ 25% on the far left for the highest Doppler precision stars. As long as high precision Doppler stars are an unbiased sample of Sun-like stars this indicates that at least $\sim$ 25% of Sun-like stars possess planets.

However, as we decrease the σ'_v threshold to consider only the highest precision stars, we are using fewer target stars to infer the fraction. Thus, the error bars on the estimates increase from $15 \pm 5\%$ on the right to $25 \pm 15\%$ on the left.

4. Fitting for and extrapolating trends

We use extrapolation to estimate the fraction of stars with planets within regions of mass-period parameter space larger than the less-biased sample (Fig. 1). Since there is no accepted theoretical model for the functional form for the mass or period distribution functions, we make simple linear fits to the histograms in log $M\sin i$ and log P and we fit conventional power laws to the histograms in $M\sin i$ and P (Fig. 5).

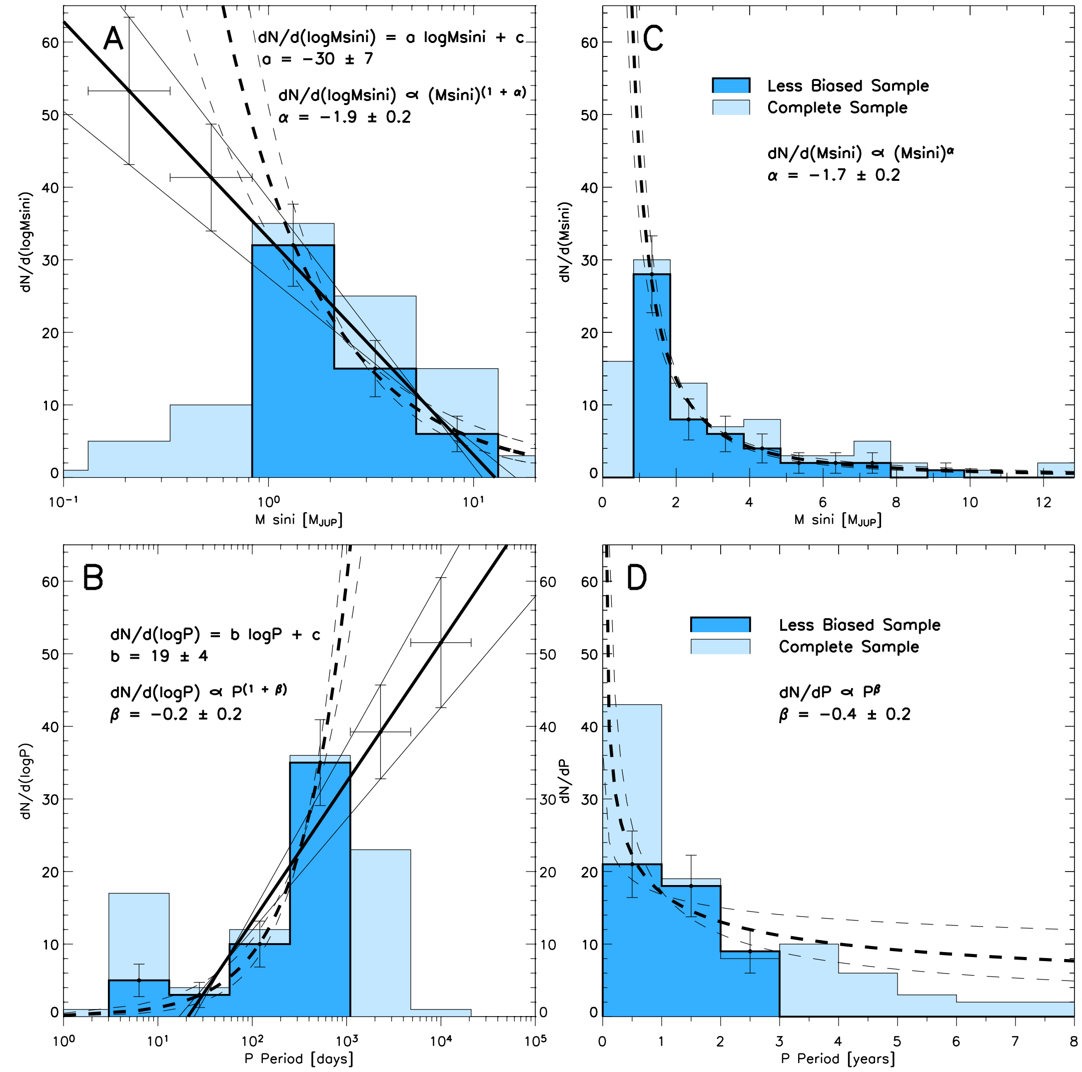

Fig. 5.— Trends that we extrapolate to lower mass and longer periods. Histograms of planet mass (top) and period (bottom) for the less-biased sample of exoplanets within the thin solid rectangle of Fig. 1, compared to all exoplanet detections. Both log (left) and linear (right) versions are shown. Histograms in log $M\sin i$ and log P (panels A & B) are fitted with the linear functional forms $dN/d(logMsini) = a\ logMsini + c$ and $dN/d(logP) = b\ logP + d$ where the best fits to the histograms yield $a = -30 \pm 7$, and $b = 19 \pm 4$. The linearly binned histograms of $M\sin i$ and P (panels C & D) are fitted to the functional forms: $dN/d(M\sin i) \propto (M\sin i)^{\alpha}$ and $dN/dP \propto P^{\beta}$ respectively and we find $\alpha = -1.7 \pm 0.2$ and $\beta = -0.4 \pm 0.2$. To check the robustness of the fits we change variables to log $M\sin i$ and log P (effectively producing a re-binning of the data) and fit for α and β in panels A & B. We find $\alpha = -1.9 \pm 0.2$ and $\beta = -0.2 \pm 0.2$. Combining these two estimates gives our best estimates of $\alpha = -1.8 \pm 0.3$ and $\beta = -0.3^{+0.3}_{-0.4}$. In the fit of panel B, we ignore the 5 exoplanets in the smallest period bin because we are interested in the overall pattern that can be most plausibly extrapolated to larger P bins, not in the pile up associated with a poorly understood stopping mechanism at $P < 12$ days. For consistency, these 5 exoplanets have also been removed from the smallest period bin in panel D.

Fitting $dN/d(logMsini) = a\ logMsini + c$ to the histogram of log $M\sin\ i$ (Fig. 5**A**) we obtain the slope $a = -30 \pm 7$. Fitting $dN/d(logP) = b\ logP + d$ to the histogram of log P (Fig. 5**B**) we obtain the slope $b = 19 \pm 4$. These trends are shown as thick lines. The distribution of exoplanets is not flat in either log $M\sin\ i$ or log P. That is $a = -30 \pm 7$ is significantly different ($\sim 4\sigma$) from $a = 0$ and $b = 19 \pm 4$ is significantly different ($\sim 5\sigma$) from $b = 0$. Fitting $dN/d(Msini) \propto (Msini)^{\alpha}$ to the histogram of $M\sin\ i$ (Fig. 5**C**) we obtain $\alpha = -1.7 \pm 0.2$. Fitting $dN/dP \propto P^{\beta}$ to the histogram of P (Fig. 5**D**) we obtain $\beta = -0.4 \pm 0.2$.

Since $d\ (ln\ x) = dx/x$, the functional form $dN/d(Msini) \propto (Msini)^{\alpha}$ can be written as $dN/d(logM\sin\ i) \propto (M\sin\ i)^{(1+\alpha)}$ and similarly $dN/dP \propto P^{\beta}$ can be written as $dN/d(logP) \propto P^{(1+\beta)}$. We fit these functions to the histograms of log $M\sin\ i$ and log P in Fig. 5**A** & **B** where we find $\alpha = -1.9 \pm 0.2$ and $\beta = -0.2 \pm 0.2$. These values differ by $\sim 1\sigma$ from the values of α and β found in Fig. 5**C** & **D**. We attribute the differences to the different log and linear binning schemes. We combine the two values into our best estimates: $\alpha = -1.8 \pm 0.3$ and $\beta = -0.3^{+0.3}_{-0.4}$. These values are compared to other estimates in Table 1. If the distributions were flat in log $M\sin\ i$ and log P we would find respectively $\alpha \approx -1.0$ and $\beta \approx -1.0$. However, both the trend in mass (α) and in period (β) are significantly different ($\sim 2\sigma$) from flat. These slopes agree with our previous results (Table 1) and show that the evidence supporting the idea that Jupiter lies in a region of parameter space densely occupied by exoplanets, has gotten stronger in the sense that our new value $b = 19$ is larger than our previous estimate and our new values for a and α are more negative than our previous estimates.

The main differences between our results and previous results is that we obtain a steeper slope (more negative α) when we fit the funtional form $dN/d(M\sin\ i) \propto (M\sin\ i)^{\alpha}$ to the mass histogram of the data: we get $\alpha \approx -1.8$ while other analyses get $\alpha \approx -1.1$. Thus we predict more low mass planets relative to the more easily detected number of high mass planets, than do other analyses. We obtain a more shallow slope (less negative β) when we fit the funtional form $dN/dP \propto P^{\beta}$ to the period histogram of the data. We get $\beta \approx -0.3$ while other analyses get $\beta \sim -0.8$. Thus, we predict more hosts of long period planets relative to the more easily detected short period planets, than do other analyses. These differences are largely due to the differences in the way the incompleteness in the lowest mass bin and the longest period bin are accounted for. This can be seen, for example in the $\alpha = -0.7$ reported by Marcy *et al.* (2003) when no correction is made for incompleteness in the lowest mass bin. When we make no completeness correction and include the lowest mass bin, we reproduce their result.

4.1. Extrapolation Using Discrete Bins

Based on slopes a and b, we make predictions for the fraction of Sun-like stars with planets within two regions of the log mass - log period plane. The estimated populations of the two lowest mass bins based on the slope a are indicated in Fig. 5**A**. Similarly the estimated populations of the two largest period bins based on slope b are shown in Fig. 5**B**. From the less-biased region (thin rectangle in Fig. 1) we extrapolate both in $M\sin i$ and P by 1 bin. When we extrapolate in $M\sin i$ or P we are doing so at a fixed range in P or $M\sin i$ respectively. However to estimate the fraction of stars within the thick rectangle in Fig. 1 (given by the $M\sin i$ and P ranges $0.3 < M\sin i/M_{Jupiter} < 13$ and 3 days $< P <$ 13 years) we need to estimate the number of planets N_{MPext} in the bottom right-hand region (just below Jupiter, see Fig. 1). To do this we use Eq. 1 (see Fig. 6). We have $N_{lb} = 53 \pm 2$ from Fig. 1 and $N_{Mext} = 41 \pm 7$ and $N_{Pext} = 39 \pm 6$ from Fig. 5**A**&**B**. Inserting these values into,

$$\frac{N_{Mext}}{N_{lb}} \approx \frac{N_{MPext}}{N_{Pext}}, \tag{1}$$

(see Fig. 6) we obtain $N_{MPext} = 30 \pm 9$. Thus, $N_{hosts} = 163 \pm 24$ and the fraction f, of targets hosting planets is the ratio,

$$f = \frac{N_{hosts}}{N_{targets}}, \tag{2}$$

where $N_{targets} \sim 1812 \pm 103$ (Table 4). Finally we find $f = (163 \pm 24)/1812 \pm 103) \approx 9 \pm 1\%$. Thus, using the slopes a and b to extrapolate one bin into lower masses and longer periods (thick solid rectangle in Fig. 1) we find that $9 \pm 1\%$ of the targeted Sun-like stars have planets.

Similarly and more speculatively we estimate the fraction contained within the thick dashed rectangle in Fig. 1 by extrapolating one more bin in mass (over the same range in period as the previous mass extrapolation) and by extrapolating one more bin in period (over the same range in mass as the previous period extrapolation). The analogous numbers are $N_{lb} = 53 \pm 2$, $N_{Mext} = (41 \pm 7) + (53 \pm 10)$, $N_{Pext} = (39 \pm 6) + (52 \pm 9)$ (where the sum of the 2 sets of numbers is the sum of the 2 separate bins). Equation 1 then yields $N_{MPext} = 161 \pm 50$. Summing the four contributions as before yields $N_{hosts} = 399 \pm 84$ and thus $f = (399 \pm 84)/1812 \pm 103) \approx 22 \pm 5\%$. Thus, we estimate that $22 \pm 5\%$ of the monitored Sun-like stars have planets in the larger region ($0.1 < M\sin i/M_{Jupiter} < 13$ and 3 days $< P <$ 60 years) that encompasses both Jupiter and Saturn. This larger region is less than 20% of the area of the log mass - log period plane occupied by our planetary system.

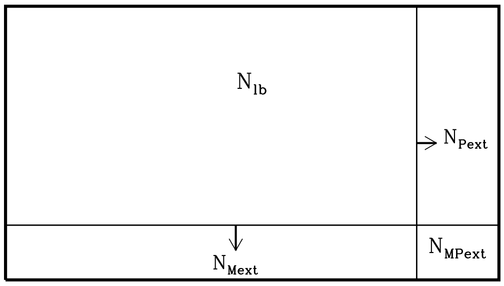

Fig. 6.— The simple method of extrapolation used to predict planet numbers in under-sampled regions of the log mass - log period plane. The known value of N_{lb} (number of planet hosts in the less-biased region of Fig. 1) and extrapolations based on the slopes a and b were used to derive N_{Mext} and N_{Pext}. We then use Eq. 1 to estimate N_{MPext}. Our estimate of the total number of planet hosts is then $N_{hosts} = N_{lb} + N_{Mext} + N_{Pext} + N_{MPext}$.

Table 1: Best-fit trends in fits to mass and period histograms and comparison with other analyses

Source	α	β	a	b
this paper, Fig. 4 A,B[j]	–	–	-30 ± 7	19 ± 4
" " Fig. 4 C,D[k]	-1.7 ± 0.2	-0.4 ± 0.2	–	–
" " Fig. 4 A,B[l]	-1.9 ± 0.2	-0.2 ± 0.2	–	–
" "combined Fig. 4 A,B,C,D	-1.8 ± 0.3	$-0.3^{+0.3}_{-0.4}$	–	–
Marcy, Butler, Fischer & Vogt 2003	-0.7[m]	–	–	–
Lineweaver, Grether & Hidas 2003	-1.6 ± 0.2	–	–	13 ± 4
Lineweaver & Grether 2002	-1.5 ± 0.2	–	-24 ± 4	12 ± 3
Tabachnik & Tremaine 2002	-1.11 ± 0.10	-0.73 ± 0.06	–	–
Stepinski & Black 2001	-1.15 ± 0.01	-0.98 ± 0.01	–	–
Jorissen, Mayor & Udry 2001	~ -1	–	–	–
Zucker & Mazeh 2001	~ -1	–	–	–

[j] best-fit of $dN/d(logM\sin\ i) = a\ logM\sin\ i + c$ to histogram of log $M\sin\ i$ and best-fit of $dN/d(logP) = b\ logP + d$ to histogram of log P

[k] best-fit of $dN/d(M\sin i) \propto (M\sin i)^{\alpha}$ to histogram of $M\sin i$ and best-fit of $dN/dP \propto P^{\beta}$ to histogram of P

[l] best-fit of $dN/d(logM\sin\ i) \propto (M\sin\ i)^{1+\alpha}$ to histogram of log $M\sin\ i$ and best-fit of $dN/d(logP) \propto P^{1+\beta}$ to histogram of log P

[m] including the lowest mass bin with no completeness corrections

4.2. Extrapolation Using a Differential Method

Instead of extrapolating one or two discrete bins, we can generalize to a more flexible differential method. For example, we can use the power law functional form to integrate a differential fraction within an arbitrary range of $M\sin\, i$ and P (Tabachnik & Tremaine 2002),

$$df = c(M\sin\, i)^{\alpha} P^{\beta}\; d(M\sin\, i) dP. \tag{3}$$

We find the normalization c by inserting the known values from the less-biased area: $f = (53 \pm 2/(1812 \pm 103)) = 2.9 \pm 0.2\%$, $\alpha = -1.8 \pm 0.3$ and $\beta = -0.3^{+0.3}_{-0.4}$ (Table 1). We integrate between the boundaries of the less-biased rectangle (3 days $< P <$ 3 years and $0.84 < \frac{M\sin\, i}{M_{Jupiter}} < 13$) and solve for the normalization. We find $c = 8.6^{+40.7}_{-5.8} \times 10^{-5}$. Under the assumption that this same c, α and β hold over larger regions of parameter space, we can integrate over the larger region and solve for f. For the thick solid and thick dashed regions shown in Fig. 1, we find $f = 20^{+23}_{-10}\%$ and $100^{+0}_{-69}\%$ respectively. Using this differential method with the slopes a and b we find values nearly identical with the discrete bin extrapolations based on a and b: $f = 9 \pm 1\%$ and $22 \pm 5\%$ respectively.

The power-law based estimates are consistent with the lower linear ($a-$ and $b-$based) estimates in the sense that the large error bars on the power-law estimates overlap with the lower estimates of the linear method. The reason the power law fits yield larger fractions can be seen in Fig 5 **A** &**B**. In the low mass and large period regions, the dashed curves are higher than the solid lines. Without further data from less massive planets and larger periods we interpret this difference as an uncertainty associated with the inability of the data to prefer one of the two simple functional forms fit here. The power law fits are marginally better fits to the data. We interpret the lower values from the linear fits as conservative lower limits to the fraction f.

4.3. Fractions in Velocity-Period Parameter Space

The fractions discussed in Sections 2 & 3 are based on exoplanet detections constrained by the Doppler technique to a trapezoidal region of the log mass - log period plane defined by a minimal velocity K_{min}. The fractions discussed in the previous two sections are based on rectangular regions of the log mass - log period plane. In the lower panel of Fig. 3, the fractions plotted as a function of duration come from a trapezoidal region (not a rectangular region) of the $M\sin\, i - P$ plane. Hence, a fit of $dN/dP \propto P^{\beta'}$ to the P histogram of detected planet hosts, will produce a value of β' slightly different from the β of Fig. 5**D**. We find $\beta' = -0.5 \pm 0.2$ (while $\beta = -0.4 \pm 0.2$). Such a difference is expected since the number of planets at large P will decrease more steeply in the trapezoidal region, because as P increases the "Detected" regions becomes narrower than a region defined by a constant $M\sin\,i$. The curve shown in the lower panel of Fig. 3 is the integral of $P^{\beta'}$. It is normalized to the last two bins on the right. In plotting this integral we are assuming that a survey of duration P_s has observed a large fraction of its target stars with Doppler-detectable exoplanets of period $P \lesssim P_s$.

An interesting consistency check of the relationships between our extracted values of α, β and K_{min} and the limits of integration for Eq. 3 is described in Appendix B.

4.4. Jupiter-like Planets

The fraction of Sun-like stars possessing Jupiter-like planets is important since Jupiter is the dominant orbiting body in our Solar System and had the most influence on how our planetary system formed. Exoplanets with Jupiter-like periods and Jupiter-like masses are on the edge of the detectable region of parameter space. The cross-hatched square in Fig. 1 is our Jupiter-like region defined by orbital periods between the period of the asteroid belt and Saturn with masses in the range $M_{Saturn} < M\sin\,i < 3M_{Jupiter}$. We estimate the fraction of Sun-like stars hosting planets in this region using the differential method based on our best-fit values of a and b. Integrating between the limits of the cross-hatched area we obtain $f = 5 \pm 2\%$ of Sun-like stars possess such a Jupiter-like planet. Using the differential method based on our best-fit values for α and β we obtain $f = 28^{+62}_{-21}\%$. As with the different f values resulting from these two methods in Sec. 4.2, these are consistent with each other and reflect the different functional forms used.

Table 2: Fraction Comparison (see Fig. 7)

Source	$M\sin\,i$ range	Period range	Fraction	Our Fraction[a]	Our Fraction[b]
TT[c] 2002	1 - 10 M_{Jup}	2 days - 10 yrs	4%	$4 \pm 1\%$	$6^{+3}_{-2}\%$
Liu *et al.* 2002	0.25-15 M_{Jup}	0-8 yrs ($<$ 4 AU)	6%	$9 \pm 3\%$	$18^{+19}_{-7}\%$
Armitage *et al.* 2002	0.3-10 M_{Jup}	0.03-11.2 yrs[d]	15%	$9 \pm 3\%$	$19^{+22}_{-10}\%$
TT[c] 2002	0.003-10 M_{Jup}	2 days - 10 yrs	18%	$45^{+19}_{-15}\%$	100%

[a] differential method based on slopes a and b
[b] differential method based on powers α and β
[c] Tabachnik & Tremaine (2002)
[d] corresponds to 0.1 - 5 AU

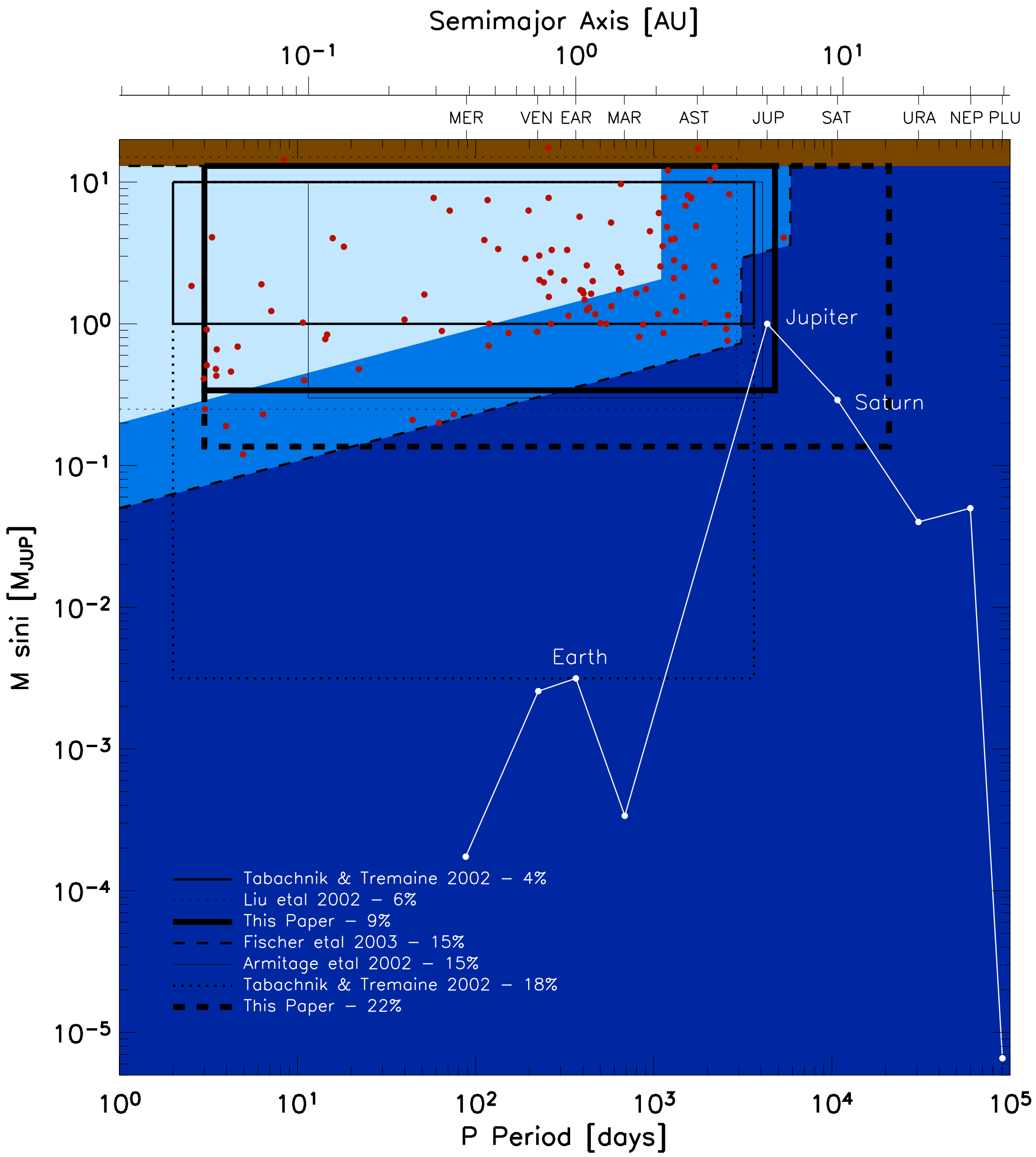

Fig. 7.— We estimate that at least 9% of Sun-like stars have planets within the thick solid rectangle and at least 22% have planets within the thick dashed rectangle (Sec. 4.1). We find larger values using a differential method based on power-law fits (Sec. 4.2). Here, we compare our results with other published estimates of the fraction of stars with planets. The papers, estimates and the regions of the log $M\sin i$ - log P plane associated with these estimates are indicated. Our results are consistent with, but generally higher than, previous work (Table 2).

5. Summary and Discussion

We have analyzed the results of eight Doppler surveys to help answer the question: 'What fraction of stars have planets?'. We use the number of targets and the number of detected planet hosts to estimate the fraction of stars with planets. Quantitatively following the consistent increase of this fraction is an important new way to substantiate both current and future estimates of this fraction. We show how the naive fraction of $\sim 5\%$ increases to $\sim 11\%$ when only long-duration targets are included. We extend the work of Fischer *et al.* (2003) by plotting the fraction as a function of excess dispersion and show how this $\sim 11\%$ increases to $\sim 25\%$ when only long-duration Lick targets with the lowest excess in radial velocity dispersion are considered.

We have identified trends in the exoplanet data based on a less-biased sub-sample. We find stronger support than found previously for the idea that Jupiter -like planets are common in planetary systems (Table 1). We estimate the fraction of Sun-like stars hosting planets in a well-defined Jupiter-like region to be $\sim 5\%$.

We have extrapolated these trends into unsampled or undersampled regions of the log mass - log period plane. We find at least $\sim 9\%$ of target stars will be found to host an exoplanet within the thick rectangle of Fig. 1 and that more speculatively at least $\sim 22\%$ of target stars will be found to host an exoplanet within the larger thick dashed rectangle. Our results for the fraction of Sun-like stars with planets are consistent with but are, in general, larger than previous estimates (Table 2, Fig. 7).

The largest uncertainty in this analysis is that we may be extrapolating trends derived from a small region of log mass - log period into regions of parameter space in which the trends are slightly or substantially different. This uncertainty is why we did not extrapolate beyond the dashed region in Fig. 1 that contains from our Solar System, only Jupiter and Saturn.

It is sometimes implicitly assumed that most planetary systems will be like ours and that Earth-like planets will be common in the Universe. However, as we descend in scale from galaxy, to star, to planetary system to terrestrial planet we run more of a risk of self-selection. That is, the factors that are responsible for our origin may have selected a non-typical location. Thus, answering the question "What fraction of stars have planets?" must rely on the continued analysis of the statistical distributions of exoplanets detected by the increasingly precise and ground-breaking Doppler surveys.

The hypothesis that $\sim 100\%$ of stars have planets is consistent with both the observed exoplanet data which probes only the high-mass, close-orbitting exoplanets and with the observed frequency of circumstellar disks in both single and binary stars. The observed fractions f that we have derived from current exoplanet data are lower limits that are consistent with a true fraction of stars with planets f_t, in the range $0.25 \lesssim f_t \lesssim 1$. If the fraction of Sun-like stars that possess planets is representative of all stars, our result means that out of the ~ 300 billion stars in our Galaxy there are between ~ 75 and ~ 300 billion planetary systems.

6. Acknowledgements

CHL acknowledges a research fellowship from the Australian Research Council. We thank Stephane Udry, Debra Fischer, Andrew Cumming and Sylvain Korzennik for help with many details of the exoplanet surveys. We thank Luis Tenorio for statistical counseling.

7. Appendix A: Estimating Target List Overlap

To compute the total number of target stars being monitored (Fig. 2) we need to avoid double-counting. Target lists of six of the eight surveys considered here have been published, thus allowing the overlap in the FGK targets of these six to be eliminated by comparing target star names (Table 3). The total number monitored by these six surveys is then known ($N_{known} \approx 1124$). The total number of targets monitored by the Coralie and Elodie surveys is known ($N_C = 1100, N_E = 350$) but without published target lists the extent of overlap with the other surveys can only be estimated. We do this by using the statistics of duplicate detections. Let the total number of exoplanet hosts discovered or confirmed by Elodie and Coralie be $N_{Ehosts} = 16$ and $N_{Chosts} = 33$ respectively. The number of these planet hosts that were discovered or confirmed by any of the other 6 surveys are $N_{Eoverlap} = 8$ and $N_{Coverlap} = 10$. Since there is no overlap between the Coralie and Elodie surveys (Udry private communication), these two sets of planet hosts are mutually exclusive. The fractional overlap of detections/confirmations is $g_E = N_{Eoverlap}/N_{Ehosts} = 8/16 = 0.50$ and $g_C = N_{Coverlap}/N_{Chosts} = 10/33 = 0.30$ (Tables 3 & 4). We use these fractional detection overlaps as estimates of the fractional target list overlaps. Thus, we estimate the total number of monitored targets (excluding overlap) as:

$$N_{total} \approx N_{known} + (1 - g_C)\ N_C + (1 - g_E)\ N_E. \tag{4}$$

However, estimated this way, N_{total} includes target stars that were found to be single line spectroscopic binaries (SB1). Exoplanet detection is difficult in such systems so we correct for this by substracting the estimated fractions of SB1's in the various surveys (between 3% and 9%, see Tables 3 & 4). After this last step we find that $N_{total} \approx 1812 \pm 103$. The time

Table 3: Doppler Surveys[a] - Targets

	AAT	Lick	Keck	Coralie	Elodie	AFOE	CES	McDonald
Targets[b]	204	360	600[c]	1100[d]	350	146	37	33
FGK IV, V Targets	198	360	443	1100	350	136	32	33
List Published	Y	Y[e]	Y[e]	N	N	Y[f]	Y	Y
Year Started	1998	1987	1996	1998	1994	1995	1992	1987
Spectral Types	FGKM	F7-K0	F7-M5	F8-M0	FGK	FGK	F8-M5	FGK
Apparent V	$\lesssim 7.5$	$\lesssim 7.5$	$\lesssim 11$	<9 (for 80%)	<7.65	$\lesssim 7.5$	$\lesssim 6$	$\lesssim 6$
Log(R'_{HK})[g]	$\lesssim -4.5$	$\lesssim -4.5$	$\lesssim -4.5$					
vsini (km/s)[h]				<4[d]	<5	<8		
Detected SB1's % [i]	8.8	3.3	3.3	8.7	3.5	3.5	8.1	3.5
Planet Host Detections[j]	18	18	26	33	16	6	2	3
References[k]	1	2,9,10	3,9,11	4	5,12	6	7	8

[a] AAT - Anglo Australian Observatory, Anglo Australian Telescope, UCLES Spectrograph
Lick - Lick Observatory, Hamilton Spectrograph
Keck - Keck Observatory, HIRES Spectrograph
Coralie - European Southern Observatory, Euler Swiss Telescope, Coralie Spectrograph
Elodie - Haute Provence Observatory, Elodie Spectrograph
AFOE - Whipple Observatory, AFOE Spectrograph
CES - European Southern Observatory, CAT Telescope, CES Spectrograph
McDonald - McDonald Observatory, Coudé Spectrograph
[b] refers to all target stars in a survey
[c] another 450 stars have one or more observations but have been dropped for various reasons (Butler *et al.* 2002)
[d] there are 550 faster rotators in a lower priority target list (Udry 2003, private communication)
[e] combined Lick/Keck (889 stars)
[f] Korzennik 2003 private communication
[g] the fractional CaII H and K flux corrected for the photospheric flux (see Noyes 1984 and Saar *et al.* 1998)
[h] projected rotational velocity
[i] SB1: single-lined spectroscopic binaries.The AAT survey (Jones *et al.* 2002) finds 18 out of 204 target stars are SB1. (Nidever *et al.* 2002) mentions that 29 out of 889 Lick and Keck target stars are SB1. (Endl *et al.* 2002) finds 3 out of 37. The original Lick survey has 5 out of 74 (Cumming *et al.* 1999). We estimate the fraction of SB1's for the other surveys by noting a survey as either northern or southern hemipshere and taking an average of the known surveys in that hemisphere to estimate the fraction of SB1's. SB2's have been eliminated from the target lists. All surveys exclude binary stars when the angular separation is less than $2''$.
[j] including confirmations but excluding confirmations of hosts that were known to have a planet prior to the start of observation. Thus several hosts of Lick and Elodie that are monitored by Keck and Coralie to increase phase coverage are only included in the Lick and Elodie numbers.
[k] 1) Jones *et al.* 2002 2) Fischer *et al.* 2003 3) Butler *et al.* 2002 4) Udry *et al.* 2000. The six newest hosts discovered by Coralie and included here were announced during the XIXth IAP Colloquium "Extrasolar Planets: Today and Tomorrow", Paris, June 2003 and are in preparation: Mayor *et al.* 2003b, Zucker *et al.* 2003 and Udry *et al.* 2003. 5) Sivan *et al.* 2000 6) Korzennik 2003, private comm. 7) Endl *et al.* 2002 8) Cochran & Hatzes 1993 9) Nidever *et al.* 2002 10) Fischer *et al.* 1999 11) Vogt *et al.* 2000 12) Mayor *et al.* website 2003a

dependence of N_{total} (Fig. 2, top panel) is taken from the time dependence of N_{known} and from N_C and N_E (Table 3) using the approximation that g_C and g_E are constants. Our estimates of the error associated with this procedure are indicated by the error bars on each bin in Figs. 2 & 3.

8. Appendix B: $M\sin i - P$ to $K - P$ parameter space

We can use the fraction $\approx 11\%$ at a duration of 15 years in the lower panel of Fig. 3 as well as the known quantities P_{max}, K_{min} to perform an interesting consistency check of the relationships between them and the extracted values of α and β. Assuming $M_{host} \approx M_\odot >> M$ and low eccentricity, the relation between $M\sin i$ and the semiamplitude K of radial velocity is

$$K \approx A\frac{M\sin i}{P^{1/3}}, \tag{5}$$

where $A = 200$ when K is expressed in units of m/s, $M\sin i$ in M_J and P in days. We are only interested in exoplanets with $M\sin i < 13M_{Jupiter}$ ($\sim$ the deuterium burning limit). Using Eq. 5 to change variables from $M\sin i$ to K we can write,

$$df = c(M\sin i)^\alpha P^\beta \; d(M\sin i) \; dP \tag{6}$$

(Eq. 3) as,

$$df = \frac{c}{A^{(1+\alpha)}} K^\alpha P^{\beta+\frac{1}{3}(1+\alpha)} \; dK \; dP. \tag{7}$$

Under the assumption that $M\sin i$ is uncorrelated with P and that α and β are approximately constant within the region of interest, we can integrate between arbitrary limits. The upper limit on K depends on period: $K_{max}(P) = 13\ A/P^{1/3}$ from Eq. 5. We then have,

$$f = \frac{c}{(1+\alpha)A^{(1+\alpha)}} \left[(13\ A)^{(1+\alpha)} \int_0^{P_{max}} P^\beta dP - K_{min}^{(1+\alpha)} \int_0^{P_{max}} P^{\beta+\frac{1}{3}(1+\alpha)} dP \right] \tag{8}$$

Using our best estimates $\alpha = -1.8$, $\beta = -0.3$, and $c = 8.6 \times 10^{-5}$ and $f = 0.106$ from the lower panel in Fig. 3 at 15 years, we integrate Eq. 8 and solve for K_{min}. We find $K_{min} = 16^{+6}_{-4}$ m/s. The weighted average internal error of the original Lick and McDonald observing programs that contain the 85 stars in these two bins is $\sigma = 6 - 10$ m/s. The minimum signal to noise of exoplanet detections is ~ 3 so we expect $K_{min} \approx 3\ \sigma$ which is indeed the case.

Table 4: Doppler Surveys - Cumulative Numbers of FGK IV-V Targets as a Function of Time

Survey(σ m/s)[a]	1988[b]	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Lick $10-15$	44	56	62	64	66	67	73	–	–	–	–	–	–	–	–	–
Lick $3-5$	–	–	–	–	–	–	–	74	74	74	265	360	360	360	360	360
McDonald $10-20$	22	33	33	–	–	–	–	–	–	–	–	–	–	–	–	–
McDonald $5-10$	–	–	–	33	33	33	33	33	33	33	33	–	–	–	–	–
McDonald ~ 3	–	–	–	–	–	–	–	–	–	–	–	33	33	33	33	33
CES $8-15$	0	0	0	0	0	29	29	29	29	32	32	32	32	32	32	32
Elodie ~ 10	0	0	0	0	0	0	0	142	142	159	315	350	350	350	350	350
AFOE ~ 10	0	0	0	0	0	0	0	23	90	113	136	136	136	136	136	136
Keck $2-5$	0	0	0	0	0	0	0	0	8	179	270	326	360	399	443	443
AAT ~ 3	0	0	0	0	0	0	0	0	0	0	66	198	198	198	198	198
Coralie ~ 3	0	0	0	0	0	0	0	0	0	0	33	700	1000	1100	1100	1100
Cumulative	66	89	95	97	99	129	135	301	376	590	1150	2135	2469	2608	2652	2652
no overlap[c]	63	85	87	87	89	111	111	210	263	413	790	1465	1688	1782	1812	1812
non-cumulative	66	23	6	2	2	30	6	166	75	214	560	985	334	139	44	0
no overlap[c]	63	22	2	0	2	22	0	99	53	150	377	675	223	94	30	0

[a] internal error also known as instrument sensitivity

[b] binning is from July to June with January at the center of bin.

[c] values have been corrected for overlapping target lists and for estimated numbers of single line spectroscopic binaries (SB1).

REFERENCES

Armitage, P.J., Livio, M., Lubow, S.H. & Pringle, J.E. 2002, MNRAS, 334, 248

Butler, P.R., Marcy, G.W., Vogt, S.S.,Fischer, D.A., Henry, G.W., Laughlin, G. & Wright, J.T. 2002, ApJ, 582, 455

Cochran, W.D. & Hatzes, A.P. 1993, in the proceedings "Planets around Pulsars", ASP Conf. Series Vol. 36

Cumming, A., Marcy, G.W. & Butler, P.R. 1999, ApJ, 526, 890

Cumming, A., Marcy, G.W., Butler, R.P. & Vogt, S.S. 2003, astro-ph/0209199 ASP Conf. Ser.

Endl, M., K*ü*rster, M., Els, M., Hatzes, A.P., Cochran, W.D., Dennerl, K. & Doebereiner, S. 2002, A&A, 392, 671

Fischer, D.A., Marcy, G.W., Butler, P.R., Vogt, S.S. & Apps, K. 1999, PASP, 111, 50

Fischer, D.A., Butler, P.R., Marcy, G.W., Vogt, S.S. & Henry, G.W. 2003, ApJ, 590, 1081

Jones, H.R.A., Butler, P.R., Marcy, G.W., Tinney, C.G., Penny, A.J., McCarthy, C. & Carter, B.D. 2002, MNRAS, submitted, astro-ph/0206216

Jorrisen, A., Mayor, M., & Udry, S. 2001, A&A in press, astroph/0105301

Lineweaver, C.H. & Grether, D. 2002, Astrobiology, 2, 3, 325

Lineweaver, C.H., Grether, D. & Hidas, M. 2003, in the proceedings "Scientific Frontiers in Research on Extrasolar Planets" ASP Conf. Ser. Vol. 294, edt Deming, D. & Seager, S., p 161, astro-ph/0209382

Liu, M.C., Fischer, D.A., Graham, J.R., Lloyd, J.P., Marcy, G.W. & Butler, P.R. 2002, ApJ, 571, 519

Marcy, G.W. & Butler P.R. 2000, PASP, 112, 137

Marcy, G.W. Butler, P.R., Fischer, D.A. & Vogt, S.S. 2003, ASP Conf. Ser. Vol. 294, edt. Deming, D. & Seager, S.

Mayor, M. & Queloz, D. 1995, Nature, 378, 355

Mayor, M. *et al.* 2003a, http://obswww.unige.ch/ñaef/planet/geneva_planets.html

Mayor, M. *et al.* 2003b, in preparation

Nidever, D.L., Marcy, G.W., Butler, P.R., Fischer, D.A. & Vogt, S.S. 2002, ApJSS, 141, 503

Noyes, R.W., Hartmann, L.W., Baliunas, S.L., Duncan, D.K. Vaughan, A.H. 1984, ApJ, 279, 763

Saar, S.H., Butler, P.R. & Marcy, G.W. 1998, ApJ, 498, L153

Sivan, J.P., Mayor, M., Naef, D., Queloz, D., Udry, S., Perrier-Bellet, C. & Beuzit, J.L. 2000, ASP Conf. Ser. Vol xxx, 2000

Stepinski, T.F. & Black, D.C. 2000, A&A, 356, 903

Tabachnik, S. & Tremaine, S. 2002, MNRAS, 335, 151

Udry, S., Mayor, M., Naef, D., Pepe, F., Queloz, D., Santos, N.C., Burnet, M., Confino, B. & Melo, C. 2000, A&A, 356, 590

Udry, S. *et al.* 2003b, in preparation

Zucker, S. & Mazeh, T. 2001, ApJ, 562, 1038

Zucker, S. *et al.* 2003, in preparation

This preprint was prepared with the AAS LaTeX macros v5.0.